1-2018

How Monitoring Influences Trust: A Tale of Two Faces

Maurice E. Schweitzer
University of Pennsylvania

Teck-Hua Ho

Xing Zhang

Recommended Citation

Schweitzer, M. E., Ho, T., & Zhang, X. (2018). How Monitoring Influences Trust: A Tale of Two Faces. *Management Science, 64* (1), 253-270. http://dx.doi.org/10.1287/mnsc.2016.2586

How Monitoring Influences Trust: A Tale of Two Faces

Abstract
Monitoring changes the behavior of those who are monitored and those who monitor others. We studied behavior under different monitoring regimes in repeated trust games. We found that trustees behaved opportunistically when they anticipated monitoring—they were compliant when they knew in advance that they would be monitored, but exploited trustors when they knew in advance that they would not be monitored. Interestingly, trustors failed to anticipate how strategically their counterparts would behave. Trustors misattributed the strategic, compliant behavior they observed as signals of trustees' trustworthiness. As a result, trustors misplaced their trust when they were unable to monitor their counterparts. We discuss the managerial implications of our results for designing and implementing monitoring systems.

Keywords
trust, compliance, monitoring, strategic behavior

How Monitoring Influences Trust: A Tale of Two Faces

Maurice E. Schweitzer
Wharton School, University of Pennsylvania
Philadelphia, PA 19104
Phone/Fax: 215.898.4776/3664
E-mail: Schweitzer@wharton.upenn.edu

Teck H. Ho
Haas School of Business
University of California at Berkeley
Berkeley, CA 94720
Phone/Fax: 510.643.4272/1420
E-mail: hoteck@haas.berkeley.edu

How Monitoring Influences Trust: A Tale of Two Faces

ABSTRACT

Organizations operate more effectively when managers trust their employees. In many cases, however, managers and their employees have divergent interests. One common managerial approach to address the problem of misaligned incentives involves monitoring employee behavior. In this paper, we investigate how monitoring changes the behavior of both those who are monitored and those who monitor others. We paired participants in a repeated trust game with a stochastic ending in which we manipulated both the frequency of monitoring and whether or not monitoring was anticipated. When trustees who were monitored could anticipate monitoring, they engaged in strategic behavior; trustees chose *self-interested* actions when they anticipated that they would *not* be monitored, but chose *altruistic* actions when they anticipated that they would be monitored. Trustors, however, failed to appreciate how strategically their counterparts would act. Though trustors were more trusting when they could monitor their counterpart than when they could not, they were still much too trusting when their inability to monitor their counterpart was anticipated. We discuss managerial implications of these results for designing and using monitoring systems.

I. INTRODUCTION

Trust plays a central role in individual relationships, in organizations, and in communities (Lewicki, Tomlinson & Gillespie, 2007). Trust promotes cooperation (Pillutla, Malhotra & Murnighan, 2003) and broadly impacts workplace behaviors (Dirks & Ferrin, 2001; Kim, Dirks, Cooper & Ferrin, 2006). Managers who develop trusting relationships with their employees are more effective and able to develop higher performance teams.

Many managers, however, may misplace their trust. For example, managers frequently trust their employees to use their work hours to accomplish organizational goals. In a recent survey, however, a majority of employees (63%) reported that they access personal e-mail at work and more than a third of employees (34%) reported that they spend at least an hour of each workday on personal internet use (Personal Internet Use Survey, 2008). These findings suggest that managerial trust may, at least some of the time, be misplaced.

Managers face a constant challenge with respect to trusting their employees. The kernel of this problem has been studied in economics using the principal-agent framework (Townsend, 1979; Mookherjee & Png, 1989). In the classic formulation of the principal-agent problem, principals (or managers) observe the outcome of their employees' efforts (i.e., outputs), but not their efforts themselves (i.e., inputs). Outcomes, however, result from a combination of employees' efforts and uncontrolled situational factors (e.g., market demand). The essence of the principle-agent problem is that managers cannot directly attribute outcomes to employee actions because managers are unable to monitor what the agent does (Grossman & Hart, 1983). Furthermore, since effort is costly for employees, managers cannot assume that employees always contribute their maximal efforts; in fact, principal-agent theory predicts that employees frequently will not exert their maximal effort.

In practice, of course, managers are able to at least partially monitor their employees. In an American Management Association survey, 66% of managers reported that they monitor the internet connections of their employees. Almost half of the companies surveyed reported that they use video monitoring, and nearly half (43%) monitor e-mail use and use other detailed tracking methods, such as tracking the keystrokes of their employees (45%). Monitoring systems, however, are expensive and their effectiveness is limited by a scare resource: managerial attention. As a result, managers can observe only a limited sample of employee behavior. For example, in most call centers supervisors listen to only a small fraction of the calls handled by the operators they oversee.

In this paper, we describe the influence of different monitoring schemes on trusting and trustworthy behaviors. We examine how people change their trustworthy behavior as a function of being monitored, and we describe how people change the trust they place in others as a function of what they observe. Specifically we study experimentally the effects of frequency of monitoring and whether or not the monitoring is anticipated on manager's trusting and employee's trustworthy behaviors respectively.

II. TRUST

A substantial literature demonstrates that trust is an essential ingredient for effective organizations (Dirks & Ferrin, 2001; Kim, Dirks, Cooper & Ferrin, 2006; Kramer & Lewicki, 2010; Lewicki, McAllister, & Bies, 1998). As Robbins (2001) explains, "People in organizations need to have positive expectations that others will not act opportunistically and take advantage of them." (p. 11) Building and maintaining trust, however, represents a constant managerial challenge (Kim, Ferrin, Cooper, and Dirks, 2004; Kim, Dirks, Cooper & Ferrin, 2006; Kramer & Lewicki, 2010). In many cases, managers rely on formal methods to obtain trustworthy behavior. For example,

managers may use formal contracts (Bottom, Holloway, Miller, Mislin, & Whitford, 2006; Malhotra & Murnighan, 2002) or material incentives (Grossman & Hart, 1983) to increase desirable behavior. These formal methods however are not always efficient. Drawing on economic theory, Stiglitz (1987) argues that managers can often achieve only "second-best" outcomes when their employees are self-interested (e.g., Stiglitz, 1987).

In this paper, we investigate the influence of a theoretically important and practically relevant, but understudied managerial tool: monitoring. We investigate behavior in a laboratory experiment with participants motivated by substantial financial incentives to examine how monitoring influences trust behavior. We build on prior work to define trust as a willingness to accept vulnerability based upon positive expectations of another person (Mayer, Davis & Schoorman, 1995), and we study how trust between paired participants changes over time (Kim, Ferrin, Cooper, & Dirks, 2004; Schweitzer, Hershey, & Bradlow, 2006).

A substantial body of research has used experimental methods to deepen our understanding of trust (e.g., Malhotra, 2004; Pillutla, Malhotra & Murnighan, 2003). This stream of research describes how incentives (Lewicki, Tomlinson, & Gillespie, 2006), formal contracts (Malhotra & Murnighan, 2002), reciprocity (Pillutla, Malhotra & Murnighan, 2003), implicit beliefs (Haselhuhn, Schweitzer & Wood, 2010), and emotions (Dunn & Schweitzer, 2005; Lount, 2010) influence trust. Much of this work has studied trust using a modified version of Berg, Dickhaut, and McCabe's (1995) trust game. In this game, an individual (the trustor) starts with a pot of money (e.g., $20) and can either keep the money for herself or pass a fraction of the money to her partner. If she passes $$x$ to her partner (the trustee), the amount of money grows (e.g., triples to 3x$), and the trustee must then decide how to split the resulting sum of money between the trustor and himself (e.g., by splitting it evenly, 1.5x$ each, or by keeping the entire 3x$ to himself). In a trust game with an

initial endowment of $20 that tripled, Berg et al. (1995) found that trustors passed an average of $5.16 (out of their initial sum of $20), and trustees returned an average of $4.66 to trustors. In this case, trust did not pay. Trustors who trusted their counterpart and passed money were worse off, on average, than those who did not.

Trust game research has also identified other important phenomena, such as the importance of vulnerability in building trust. Pillutla, Malhotra, and Murnighan (2003) found that large acts that made trustors more vulnerable were reciprocated more than small, tentative acts. Similarly, Ho and Weigelt (2005) found that trustees frequently split the money evenly if trustors' intention of trusting was unambiguously altruistic in a multi-stage trust game.

Several scholars have also used repeated games to study the evolution of trust and cooperation. In these experiments, participants play the same trust game with a fixed-matching protocol (e.g., with the same partner) multiple times. This line of research has challenged conventional wisdoms about trust, such as the belief that trust is fragile—once broken, it is either very difficult or impossible to repair (Slovic, 1993; Lewicki & Wiethoff, 2000). Experimental research has found that the timing of the breach matters (Lount, Zhong, Sivanathan & Murnighan, 2008), and that penance (Gibson, Bottom, & Murnighan, 1999; Bottom, Gibson, Daniels & Murnighan, 2002), cooperative actions (e.g., Lount, Zhong, Sivanathan & Murnighan, 2006), apologies (Dirks, Kim, Ferrin, & Cooper, 2011; Kim, Ferrin, Cooper, & Dirks, 2004), promises to change (Schweitzer, Hershey, & Bradlow, 2006), and even denials (Kim, Dirks, Cooper & Ferrin, 2006) can restore trust.

Very few studies investigate strategies trustors might adopt to contend with the problem of exploited trust. Work that has studied this problem has found that formal contracts significantly increase cooperation (Bottom, Holloway, Miller, Mislin, & Whitford, 2006; Malhotra &

Murnighan, 2002). For example, Malhotra and Murnighan (2002) found that contracts promoted very high levels of cooperation—as long as the contract was in place. When the contract was removed, cooperation collapsed.

III. MONITORING

A substantial literature suggests that monitoring is likely to change how people behave within organizations. Across many social settings, people work hard to create positive impressions (Leary, 1996), and individuals pay closer attention to how they behave when others can observe them than when others cannot (Goffman, 1959). People also strategically control the information they reveal and the emotions they express to influence others (Goffman, 1959; Andrade & Ho, 2009). Frequently, people present themselves in positive ways in order to get others to like them (Schlenker & Leary, 1982). These conscious efforts facilitate social interactions and improve efficiency in organizations (Giacalone & Rosenfeld, 1990).

In addition to making actions more visible, monitoring systems may communicate expectations. For example, by monitoring employee behavior, managers convey the belief that they expect unmonitored employees to shirk (Frey, 1993). Niehoff and Moorman (1993) found that by introducing monitoring systems, employee motivation and citizenship behavior declined. In some cases, however, monitoring can lead to positive psychological consequences, such as boosting perceptions of fairness (Niehoff & Moorman 1993).

Surprisingly little work has examined the interplay between monitoring and trust. In a field setting, Alder, Noel and Ambrose (2006) examined employee reactions to having their internet use restricted and monitored. Alder et al. (2006) found that employees trusted the organization more when management had let them know in advance that their internet use would be restricted and

monitored than when they did not. High levels of trust, however, can make members of teams reluctant to monitor each other (Langfred, 2004), and a lack of monitoring can harm performance and trustworthy behavior (Schweitzer & Ho, 2005). In this paper, we report results from a careful laboratory experiment that describes how different monitoring regimes change the trust behavior of both those who are monitored and those who monitor others.

IV. HYPOTHESES

We develop two sets of hypotheses. The first set of hypotheses considers how monitoring schemes influence the decision to be trustworthy. These hypotheses focus on how monitoring changes trustee behavior. The second set of hypotheses examines how monitoring influences the decision to trust others. These hypotheses focus on how monitoring changes trustor behavior. We analyze the dynamics of both the trustee and the trustor behaviors over time.

In developing our hypotheses, we consider two dimensions of monitoring schemes: frequently and anticipation. The frequency of monitoring reflects how often trustees are monitored. The anticipation of monitoring reflects whether or not trustees know in advance that they will be monitored. This dimension of monitoring schemes reflects the notion that some monitoring schemes let employees know in advance *(anticipated monitoring)* when they will be monitored (e.g., announced visits or advanced warning of monitoring). Other monitoring schemes do not afford trustees with advance notice *(unanticipated monitoring)*.

Trustee's behavior. People react to the behavior they observe (Lewicki, Tomlinson, & Gillespie, 2006). When people observe others' trustworthy behavior, they reciprocate by becoming more trusting. Trust builds steadily over time following positive experiences (Gambetta, 1988). Conversely, when people observe untrustworthy behavior, they respond by becoming less trusting.

As a result, trust can decline quickly and sharply (Ho & Weigelt, 2005; Lewicki, Tomlinson, & Gillespie, 2006).

Consequently, we expect trustors to react to the behavior they observe, and we expect trustees to anticipate this. Although trustees can reap short-term gains by engaging in untrustworthy behavior, their long-term profitability is harmed in repeated interactions when trustees do not trust them. Hence, we expect monitoring to change trustee behavior. We expect trustees to anticipate how trustors will respond to the trustworthy and untrustworthy behavior they observe. As a result, we expect frequent monitoring to increase trustee's trustworthy behavior. For trustees, detected untrustworthy behavior is costly, so frequent monitoring increases the expected cost of engaging in untrustworthy behavior.

H1: The frequency of monitoring is positively related to trustworthy behavior.

When trustees know in advance whether or not their behavior will be monitored, the consequences of engaging in untrustworthy behavior are clear. We expect trustees to be very likely to choose trustworthy actions when they anticipate that their behavior will be monitored. Conversely, we expect trustees to be very likely to choose untrustworthy actions when they anticipate that their behavior will not be monitored.

H2: Anticipated monitoring will increase trustworthy behavior in anticipated monitoring rounds.

H3: Anticipated monitoring will decrease trustworthy behavior in anticipated non-monitoring rounds

Trustor's behavior. Just as trustees react to monitoring, we expect trustors to react to the information they observe. When trustors observe trustworthy behavior, we expect trustors to trust their counterparts more in the future. When trustors observe untrustworthy behavior, we expect trustors to curtail their trust in their counterpart. Hence, we expect trustors to react to the frequency of monitoring. As a consequence, we expect trustors to be more trusting when the frequency of monitoring is high than when it is low.

For similar reasons, we expect trustors to be particularly trusting when they know that their counterpart anticipates monitoring. We expect this reciprocal relationship to facilitate the development of a common mental model (Gentner & Stevens, 1983; Senge, 1990) for trusting and trustworthy actions when monitoring is anticipated.

H4: The frequency of monitoring is positively related to trusting behavior.

H5:Anticipated monitoring will increase trusting behavior in anticipated monitoring rounds.

The more interesting question centers on how trustors decide to trust their counterpart when they cannot monitor their counterpart's actions during unmonitored rounds. Recent work suggests that the same piece of information is weighted more heavily if it arises from direct experience rather than from observation (Simonsohn, Karlsson, Loewenstein, & Ariely, 2008; Haselhuhn, Pope,

Schweitzer, & Fishman, 2012). As a result, trustors may overweight the trustworthy behavior they observe in anticipated monitoring rounds and mistakenly assume that their counterpart is trustworthy (rather than merely strategic). We expect direct experience with trustworthy behavior to disproportionately influence trustors. Consequently, trustors who observe trustworthy behavior in anticipated monitoring rounds may erroneously expect their counterpart to engage in trustworthy behavior in anticipated non-monitoring rounds.

H6: Trustors will misplace their trust and lose money in anticipated non-monitoring rounds

V. METHODS

We used a laboratory experiment to investigate the relationship between monitoring and trust. In this study, we randomly assigned pairs of participants to different monitoring conditions, and we recorded their behavior in a repeated trust game. Prior work has found that this type of behavioral measure of trust correlates closely with trust attitudes (McEvily, Radzevick, & Weber, 2012; Schweitzer, Hershey, & Bradlow, 2006). Participants were compensated by financial incentives and their monetary payoffs depended on their actions, and those of their partner, and chance.

V.1 TRUST GAME

We randomly assigned participants to assume the role of either the odd or the even player in a repeated trust game. In each round of the game, the odd player started with an endowment of five points. The odd player decided how much of the five points to keep for himself and how much to pass to his even player counterpart. The even player received a multiple of the amount the odd player passed. In this experiment, the amount the odd player passed was quadrupled. For example,

if an odd player passed 2 points (and kept 3 points) then the even player received 8 points (i.e., 4 x 2 = 8 points). The even player then decided how much to keep for herself and how much to pass back to the odd player. For example, if the odd player passed 2 points (and the even player received 8 points) and the even player returned 4 points, the odd player would earn 7 points (i.e., 3 + 4) and the even player would earn 4 points (i.e., 8 - 4). The round ended after the even player made her decision.

Each pair of participants in this experiment played the same game multiple times. Each time they played, participants were in the same role and were matched with the same partner. We summed the points players earned for each round that they played and converted their total points into cash; each point was worth 10 cents. Participants were paid in cash before they left the experiment.

V.2 DESIGN

In our experiment, we manipulated the nature of the feedback players received. Although both players made decisions in every round, they did not always learn what their counterpart chose in a particular round. We randomly assigned pairs of participants to one of four monitoring conditions from a 2 (Frequency of Monitoring) x 2 (Anticipated Monitoring) factorial design.

We randomly assigned about half of our participants to a *high frequency* monitoring condition and half to a *low frequency* monitoring condition. The chance that any one round would be a feedback round was constant within a condition and independent of whether or not the previous round was a feedback round (i.e. we use an *independent* draw to determine whether any one round was a feedback round). In the high frequency condition there was an 80% chance that an odd player would receive feedback each round (and learn what their even player counterpart chose for that

particular round). In the low frequency monitoring condition there was a 40% chance that an odd player would receive feedback each round (and learn what their even player counterpart chose for that particular round). In rounds without monitoring, odd players did not learn what their counterpart chose for that particular round until the experiment ended. In every round, across all conditions, even players always learned what their odd player counterpart passed to them before they made their decision.

We also randomly assigned participants to either an *anticipated* monitoring condition or an *unanticipated* monitoring condition. This condition varied whether or not even players knew before they made their decision if their current round was a feedback round. In the anticipated condition, even players knew before they made their decision whether or not their current round was a feedback round (in which case odd players would learn what they chose). In the unanticipated condition, even players did not know in advance whether or not their current round was a feedback round. Odd players always knew before they made their trusting decision whether or not the upcoming round was a feedback round. They also knew what their even player counterpart knew about the current round.

We randomly assigned pairs of participants to one of the four monitoring conditions from our 2 (frequency) x 2 (anticipated) design: high frequency-anticipated, low frequency-anticipated, high frequency-unanticipated, and low frequency-unanticipated. Both odd and even players knew details about their own treatment condition, but were unaware of the other three conditions. In sum, we used a 2x2 between-subject design with repeated measures. We measured both how much the trustor passed and how much the trustee returned in each round.

V.3 EXPERIMENTAL PROCEDURE

We recruited 204 participants from a large Northeastern university to participate in a one-hour laboratory study for a $10 show up fee and the chance to earn additional money. Participants completed the study in groups that ranged in size from 6 to 14. The average group size was 8. (If an odd number of participants arrived to the laboratory, one participant was randomly selected, removed from the group, paid $10, and dismissed.)

We instructed participants that they would have the opportunity to earn money in the experiment. We explained that the total amount that they earned would depend upon the decisions they made, the decisions their partner made, and upon chance. We seated participants at computer terminal separated by partitions. Participants were unable to see the screens of other participants. We randomly and anonymously paired participants with another person in their experimental session. Participants never found out who their paired partners were.

After reading the instructions, participants answered a series of comprehension check questions about the nature of the game and the nature of their experimental condition. The program returned participants to the instruction page if they made a mistake in answering any of the comprehension check questions. Most participants understood both the game instructions and their experimental condition. However, 13 participants made repeated mistakes answering the comprehension check questions. We dismissed both these participants and their partner. In total, we dismissed 11 dyads (in two dyads, both pair members made repeated mistakes answering the comprehension check questions). As a result, a total of 188 participants completed the study.

The total number of rounds participants played was randomly determined. Every dyad made 10 rounds of decisions. After the tenth round, participants had an 85 percent chance of continuing on to the eleventh round. If there was an eleventh round, there was an 85 percent chance of

advancing to the twelfth round, and so on. The 85 percent chance of advancing to the next round was independent of whether participants had advanced in previous rounds.

After the last round (i.e., the game failed to advance to the next round), we revealed to all participants their entire history and how many points they had earned in total. We paid participants $0.10 for every point that they had earned. The introductory instructions and the comprehension check questions covered both the stochastic ending and the payment scheme.

Across conditions, we measured the amounts odd players passed and even players returned. We compared the amounts participants passed and returned in monitored and non-monitored rounds, and we conducted analyses for all rounds as well as for just the first 10 rounds.

VI. RESULTS

Of the 188 participants who completed the study, just over half of them were female (57.45%). Dyads were randomly and independently assigned to each condition; a total of 54 participants completed the study in the high frequency-anticipated condition, 44 in the low frequency-anticipated condition, 42 in the high frequency-unanticipated condition, and 38 in the low frequency-unanticipated condition.

In analyzing our results, we report summary statistics as well as results from a set of logistic regression equations. In our experimental design, we used a stochastic ending. This aspect of our design enabled us to eliminate any end-game effect, but caused different dyads in our experiment to complete different numbers of rounds. For completeness, we report two sets of analyses. One set of analyses for the first 10 rounds, which every dyad experienced, and a second set of analyses for all of the rounds.

In Figures 1 and 2, we show participant behavior from two of the dyads in the study. In both figures, the closed circles represent the amounts odd players passed in each round, and the open squares represent the amounts the even player returned in each round. The decision round is depicted along the x-axis, and the asterisk indicates that a particular round was a monitoring round.

In Figure 1, we report results for a dyad in the "Anticipated 40%" condition. In this condition, each round had an independent, 40% probability of being a feedback or monitoring round, and both odd and even players knew *in advance* whether or not the odd player would observe what the even player chose for that particular round. As shown in Figure 1, the even player made a series of strategic choices. In monitoring rounds, the even player always returned at least twice what the odd player passed. In non-monitoring rounds, however, the even player always returned 0. The odd player in this dyad, however, did not observe the even player's strategic behavior in the non-monitoring rounds. During the experiment, the odd player was only able to observe the even player's behavior in the monitoring rounds (rounds 5, 8, 11, 13, 14, 15, 16, 17, 19, and 22). Hence, the odd player only observed that whenever s/he passed money, s/he received a substantial amount in return. The odd player in this dyad failed to appreciate how differently his/her counterpart would act when monitoring was not anticipated, and the odd player continued to pass substantial amounts in every non-monitoring round.

In Figure 2, we depict results for a dyad in the "Unanticipated 80%" condition. In this condition, each round had an independent, 80% probability of being a monitoring round, and though odd players knew in advance whether or not they would receive feedback for a particular round, even players did not know in advance whether or not their odd player counterpart would observe what the even player chose for that particular round. In stark contrast to the dyadic behavior in

Figure 1, both even and odd players were about equally likely to pass and return in monitoring and non-monitoring rounds.

Trustee's trustworthy behavior: In Table 1, we report the average amounts that participants passed/returned across conditions for the monitoring and non-monitoring rounds. The right panel of Table 1 (labeled as All Rounds) shows that even players returned the most in anticipated feedback rounds (7.05 points and 7.44 points) and returned the least in anticipated no-feedback rounds (3.17 points and 3.64 points). The discrepancy between what even players returned in the anticipated feedback and the anticipated non-feedback rounds was a function of two things: the amounts odd players passed (which influenced how much they were able to return) and even players' strategic behavior.

Table 2 reports the average percentages that participants passed/returned across conditions for the monitoring and non-monitoring rounds. The right panel of Table 2 (labeled as All Rounds) shows that, on average, even players returned the highest percentage of their endowment in the anticipated feedback rounds, 40% and 42% in the "Anticipated 40%" and the "Anticipated 80%" conditions. Even players returned the lowest percentage of their endowment in the anticipated no-feedback rounds, 18% and 24% on average in the "Anticipated 40%" and the "Anticipated 80%" conditions. These results suggest that even players were strategic in their returning behaviors in anticipated conditions.

As expected, when even players could not anticipate whether or not a round would be a monitoring round, they returned very similar amounts in both monitoring and non-monitoring rounds. The right panel of Table 2 shows that even players returned 32% of their endowment in both the monitoring and non-monitoring rounds in the "Unanticipated 80%" condition, and 33%

and 30% of their endowment in the monitoring and non-monitoring rounds, in the "Unanticipated 40%" condition.

Trustor's trusting behavior: The right panel of Table 1 shows that in the anticipated monitoring conditions (both "Anticipated 40%" and "Anticipated 80%" conditions), odd players passed the most in anticipated monitoring rounds (3.83 and 4.12), and the least in anticipated non-monitoring rounds (2.90 and 3.58). Recall that odd players always knew in advance whether or not they would observe trustee's behavior in a particular round; what varied was whether or not even players knew in advance whether or not their counterpart would observe their behavior in a particular round. The right panel of Table 1 also shows that odd players passed similar amounts in monitoring and non-monitoring rounds in the unanticipated monitoring conditions. Specifically, odd player passed on average 3.24 (40% unanticipated) and 3.32 (80% unanticipated) in monitoring rounds and 3.00 and 3.08 in corresponding non-monitoring rounds respectively.

Logistic Regression Analysis

To formally analyze our results, we conducted logistic regression analyses. We report results for four related models in Table 3. In these models, we created binary variables to denote whether or not odd players chose to trust their counterpart by passing 5 (their total endowment), and whether or not even players chose trustworthy behavior by returning a substantial amount of money. We defined even player's trustworthy behavior as returning half of their endowment (10) to the odd player (models 1 and 2) or as returning at least what their odd player counterpart had passed to them (5) in models 3 and 4. In models 1 and 3 we analyzed just the first 10 rounds of decisions (which every dyad experienced), and in models 2 and 4 we analyzed all of the rounds of decisions. We used these binary variables as dependent variables in our logistic regression analysis. The independent

variables were: whether or not the monitoring was anticipated (*A*), whether or not the frequency of feedback was high or low (*H*), whether or not a monitoring or feedback round was anticipated (*F(r)*A*), the percentage of trustworthy behavior in feedback rounds (*n(t)*), and whether or not trustworthy behavior was anticipated (A*n(t))).

The results at the bottom panel of Table 3 allow us to test Hypotheses 1-3. Supporting Hypothesis 1, we find that even players returned more when the frequency of monitoring was high. This is true across all four models, with significant coefficients for H. Consistent with Hypothesis 2, even players were very strategic with respect to their behavior. Even players returned much more in the anticipated feedback rounds. This is true across all four models, with significant coefficients for F(r)*A. Even players, however, returned significantly less in the anticipated conditions overall, represented by the high negative value of the coefficient for A, which was significant in all four models, confirming Hypothesis 3.

The results at the top panel of Table 3 allow us to test Hypotheses 4-6. Table 3 shows that odd players passed more to even players when the frequency of monitoring was high. Across the four models, the coefficient for H (β_H) was consistently significant, confirming Hypothesis 4. There was a significant interaction between feedback and anticipation. Odd players passed more when they knew a forthcoming round was a feedback and anticipated round. This is evidenced by the significant coefficient (β_{FA}) across four all models, supporting Hypothesis 5.

To explore whether odd players respond to past trustworthy behavior of even players we included the fraction of even players' trustworthy behavior in feedback rounds. Consistent with prior work demonstrating that people respond to positive feedback, we find that odd players passed more when trustees had been trustworthy. This is true across all four models, with significant coefficients for n(t).

Supporting Hypothesis 6, odd players failed to anticipate how strategically even players would act in anticipated no-feedback rounds. Though odd players passed less in the anticipated no-feedback rounds than they did in the anticipated feedback rounds, they were still more trusting than the return would justify (i.e. they made less money by passing). This was evidenced by the coefficient value for A, which was significant in 3 out of 4 models, and less negative than the coefficient value for A in the return behavior for the even player.

For each round in which odd players passed money to their even counterpart, we calculated their rate of return. (Recall that the amount odd players passed to their even counterpart quadrupled.) If even players returned half of what they received, odd players would earn a 100% rate of return. For example, if an odd player passed 5, her even counterpart would receive 20. If the even counterpart returned 10, the odd player would earn a 100% rate of return on the 5 she passed. If even players returned the same amount odd players passed, odd players would earn a 0% rate of return, and if even players returned no money, odd players would earn a -100% rate of return.)

When feedback was unanticipated, the rate of return was similar for both the feedback and no-feedback rounds. In the 80% feedback condition, the rates of return were 59.95% and 66.83%, respectively. In the 40% feedback condition, the rates of return were 33.86% and 34.4%, respectively.

More interestingly, the rates of return were dramatically different across the feedback and no-feedback rounds in the anticipated feedback conditions. In the 80% feedback condition, odd players earned a 73.91% rate of return in feedback rounds, but -0.6% in the no-feedback rounds. In the 40% feedback condition, odd players earned a 70.37% rate of return in feedback rounds, but -11.9% in no-feedback rounds. That is, when monitoring was anticipated, odd players earned

substantial returns in monitored rounds, but lost money when their counterpart anticipated that a lack of monitoring.

VII. **DISCUSSION**

Our work investigates how monitoring schemes influence trust behavior. We report results from a repeated trust game experiment with participants motivated by substantial financial incentives. We examine how different monitoring systems influence the behavior of both those who are monitored and those who monitor and observe their counterpart's behavior.

We demonstrate that monitoring schemes significantly influence behavior. When people are monitored frequently, they are more likely to choose trustworthy actions than when they are monitored infrequently. Similarly, people who monitor frequently are more likely to choose trusting actions. These individuals can observe whether their trusting actions are reciprocated with trustworthy behaviors.

When individuals anticipate being monitored, they behave strategically. In our experiment, individuals who knew when they would be monitored chose trustworthy actions when they would be monitored and untrustworthy actions when they would not be monitored. Prescriptively, these findings suggest that monitoring systems, if possible, should be unanticipated. When even players did not know in advance whether or not their behavior would be monitored, they were significantly more trustworthy than when they knew in advance whether or not their behavior would be monitored. This uncertainty promoted trustworthy behavior.

Individuals who observed others' behaviors chose more trusting actions in monitored than unmonitored rounds. Importantly, however, these individuals failed to anticipate how strategically their counterpart would behave. Though odd players trusted their counterparts less when they

anticipated no monitoring, their adjustments were insufficient. Odd players in these rounds passed far more money to even payers then they received in return. As a result, trusting behavior in anticipated non-monitored rounds resulted in a negative return. This mental model failure is striking because both the odd and even players were drawn from the same subject pool and randomly assigned to each role in our study.

Like contracts (Bottom et al., 2002; Bottom et al., 2006; Malhotra & Murnighan, 2002), monitoring systems attempt to promote trustworthy behavior by changing incentives. By increasing the visibility of behavior, monitoring allows trustors to punish those they observe to be untrustworthy. Our research suggests that this approach for promoting trust can have important, unintended consequences: managers may systematically misplace their trust when employees expect not to be monitored. In addition, monitoring systems harm an organizational culture by communicating distrust.

The problem of misplacing trust may be particularly important for managers. Not only do managers routinely face the challenge of trusting others, but they may also be especially ill-equipped to content with it. Perspective taking within trust relationships is difficult in general (Weber, Malhotra, & Murnighan, 2004), and high status individuals, like managers, tend to be particularly trusting (Lount & Pettit, 2012).

Our research demonstrates that in addition to consuming scarce managerial attention, monitoring systems can change the relationship between managers who monitor and employees who are being monitored. Future research should explore the full range of consequences of monitoring systems. In addition, future work should investigate other approaches for promoting trust, such as building a sense of shared identity (Shapiro, Sheppard, Cheraskin 1992).

Managers face a constant challenge with respect to trusting their employees. Our work demonstrates that managers should view the behavior they observe skeptically, consider changing the nature of the monitoring system they use, and recognize that what they cannot see may be far more indicative of trustworthiness than what they can.

REFERENCES

Aiello, J. 1993. Computer-based monitoring: Electronic surveillance and its Effects. *Journal of Applied Social Psychology,* 23, 499-507.

Alder, S., Noel, T., & Ambrose, M. (2006). Clarifying the Effects of Internet Monitoring on Job Attitudes: The Mediating Role of Employee Trust, *Information and Management*, 43, 894-903.

American Management Association (2007), *Electronic Monitoring and Surveillance Survey*, http://www.amanet.org/news/177.aspx, (accessed 21 May 2010).

Andrade E. and Ho T. (2009). Gaming Emotions in Social Interactions, *Journal of Consumer Research,*36:539-552.

Bazerman, M. H. 1994. *Judgment in Managerial Decision Making*. New York: Wiley.

Berg, J., Dickhaut, J. & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10, 122-142.

Boles, T., Croson, R., Murnighan, J. (2000). Deception and Retribution in Repeated Ultimatum Bargaining, *Organizational Behavior and Human Decision Processes*, 83, 235-259.

Bottom, W., Gibson, K., Daniels, S. & Murnighan, K. (2002). When talk is not cheap: Substantive penance and expressions of intent in rebuilding cooperation. *Organization Science*, 13, 297-513.

Bottom, W., Holloway, J., Miller, G., Mislin, A., & Whitford, A. (2006). Building a pathway to cooperation: Negotiation and social exchange between principal and agent. *Administrative Science Quarterly*, 51, 29-58.

Carr, A. (1968). Is business bluffing ethical? *Harvard Business Review*, 46, 143-153.

Dirks, K. (1999). The effects of interpersonal trust on work group performance. *Journal of Applied Psychology*. 84, 445-455.

Dirks, K. T., & Ferrin, D. L. (2001). The role of trust in organizational settings. *Organization Science, 12*, 450-467.

Dirks, K., Kim, P., Ferrin, D. & Cooper, C. (2011). Understanding the effects of substantive responses on trust following a transgression, *Organizational Behavior and Human Decision Processes*, 114, 87-103.

Dunn, J. & Schweitzer, M. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology*, 88(6), 736-748.

Ferrin, D. & Dirks, K. (2003). The use of rewards to increase and decrease trust: Mediating processes and differential effects. *Organization Science*, 14, 18-31.

Frey, B. (1993). Does monitoring increase work effort? The rivalry with trust and loyalty. *Economic Inquiry*, 4, 663-670.

Giacalone, R. & Rosenfeld, P. (1990) Effect of self-presentation on perceptions and recognition in an organizational, *Journal of Psychology*, 124, 25-38.

Gambetta, D. 1988. Can we trust trust? In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations*: 213-237. Cambridge, MA: Basil Blackwell.

Gibson, K., Bottom, W. & Murnighan, K. (1999). Once bitten: Defection and reconciliation in a cooperative enterprise, *Business Ethics Quarterly*, 9, 69-85.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Doubleday Anchor.

Gentner, D. & Stevens, A.(1983). *Mental Models*. Hillsdale, NJ: Erlbaum.

Gibson, K., Bottom, W., Murnighan, J. (1999). Once bitten: Defection and reconciliation in a cooperative enterprise. *Business Ethics Quarterly*, 9, 69-85.

Glaeser, E., Laibson, D., Scheinkman, J., & Soutter, C. (2000). Measuring trust. *The Quarterly Journal of Economics*, 115, 811-846.

Grossman, Sanford and Oliver Hart. 1983. An Analysis of the Principal-Agent Problem, *Econometrica*, 51, pp. 7-45.

Haselhuh, M., Pope, D., Schweitzer, M., Fishman, P. (2012). The Impact of Personal Experience on Behavior: Evidence from Video-Rental Fines, *Management Science*, 58, 52-61.

Haselhuhn, M., Schweitzer, M. Wood, A. (2010). How implicit beliefs influence trust recovery, *Psychological Science*, 21 (2010), pp. 645–648.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. and McElreath, R.: *In search of homo economicus: behavioral experiments in 15 small-scale societies,* Amer. Econ. Rev. 91 (2001), 73-78

Ho, T. & Weigelt, K. (2005). Trust Building Among Strangers. *Management Science*, 51, 519-530.

Huang, L. & Murnighan, J. (2010). What's in a name? Subliminally activating trusting behavior, *Organizational Behavior and Human Decision Processes*, 111, 62-70.

Kim, P. H., Dirks, K. T., Cooper, C., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence- vs. integrity-based violation. *Organizational Behavior and Human Decision Processes, 99,* 49-65.

Kim, P., Ferrin, D., Cooper, C., & Dirks, K. (2004). Removing the shadow of suspicion: The effects of apology versus denial for repairing competence-versus integrity-based trust violations. *Journal of Applied Psychology*, 89, 104-118.

Kramer, R. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology*, 50: 569-598.

Kramer, R. & Lewicki, R. (2010) Repairing and enhancing trust: Approaches to reducing organizational trust deficits, *The Academy of Management Annals*, 4, 245–277.

Kramer, R. & Tyler, T. (1996) *Trust in Organizations*. Thousand Oaks, CA: Sage.

Kruglanski, A. (1970). Attributing trustworthiness in supervisor-worker relations. *Journal of Experimental Social Psychology*, 6, 214-232.

Langfred, C. (2004). Too much of a good thing? Negative effects of high trust and individual autonomy in self-managing teams, *Academy of Management Journal*, 47, 385-399.

Leary, M. (1996). *Self-Presentation: Impression Management and Interpersonal Behavior*. Dubuque, IA: Brown & Benchmark.

Lewicki, R., McAllister, D. & Bies, R. (1998) Trust and distrust: new relationships and realities, *Academy of Management Review*, 23, 438-458.

Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. 2006. Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, 32: 991–1022.

Lewicki, R. & Wiethoff, C. (2000). Trust, trust development, and trust repair. In The *handbook of conflict resolution: Theory and Practice*. Deutsch, M. and Coleman, P. (Eds.), San Francisco: Jossey-Bass, 86-107.

Lount, R. (2010). The impact of positive mood on trust in interpersonal and intergroup interactions. *Journal of Personality and Social Psychology*. 98, 420-433.

Lount R., Zhong C., Sivanathan N., Murnighan K.(2006). Getting off on the wrong foot: The timing of the breach and the restoration of trust. *Journal of Personality and Social Psychology*. 34: 1601-1612.

Lount, R. & Pettit, N. (2012). The social context of trust: The role of status, *Organizational Behavior and Human Decision Processes*, 117, 15-23.

Malhotra, D. (2004) Trust and reciprocity decisions: The differing perspectives of trustors and trusted parties, *Organizational Behavior and Human Decision Processes*, 94, 61-73.

Malhotra, D., & Murnighan, J. K. (2002). The effects of formal and informal contracts on interpersonal trust. *Administrative Science Quarterly*, 47, 534–559.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20: 709-734.

McEvily, B., Radzevick, J. & Weber, R. (2012). Whom do you distrust and how much does it cost? An experiment on the measurement of trust, *Games and Economic Behavior*, 74, 285-298.

McKnight, D., & Chervany, N. (2002). What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. *International Journal of Electronic Commerce, 6*, 35–60.

Mookherjee, D. & Png, I. (1989). Optimal auditing, insurance, and redistribution. *The Quarterly Journal of Economics*, 104, 399-415.

Niehoff, B. P., & Moorman, R. H. 1993. Justice as a mediator of the relationship between methods of monitoring and organizational citizenship behavior. *Academy of Management Journal*, 36: 527-556.

Pillutla, M., Malhotra, D., & Murnighan, J. K. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology*, 39, 448–455.

Schlenker, B. & Leary, M. (1982). Social anxiety and self-presentation: a conceptualization and model, *Psychological Bulletin*. 92:641-69.

Robbins, S. (2001) *Organizational Behavior*, 9th Edition, Upper Saddle River, New Jersey.

Sappington, David. (1991). Incentives in Principal-Agent Relationships, *Journal of Economic Perspectives* 5, 45-66.

Schweitzer, M., Hershey, J., & Bradlow, E. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes,* 101(1), 1-19.

Schweitzer, M. & Ho, T. (2005). Trust but verify: Monitoring in interdependent relationships, in *Experimental and Behavioral Economics* (Ed. J. Morgan), 13, 87-106.

Simonsohn, U., Karlsson, N., Loewenstein G., & Ariely, D. (2008). The tree of experience in the forest of information: Overweighing experienced relative to observed information, *Games and Economic Behavior,* 62, 263-286.

Senge, P. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday.

Shapiro, D., B. H. Sheppard, and L. Cheraskin (1992), "Business on a Handshake," Negotiation Journal, 8 (4), 365-77.

Slovic, P. (1993). Perceived risk, trust, and democracy. *Risk Analysis*, 13, 675–682.

Stiglitz, Joseph E. (1987). "Principal and agent, *The New Palgrave: A Dictionary of Economics*, v. 3, pp. 966-71.

Townsend, R. (1979).Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 22, 265-293.

Weber, R., & Camerer, C. (2003). Cultural conflict and merger failure: An experimental approach. *Management Science*, 49, 400-415.

Weber, R., Camerer, C., Rottenstreich, Y., & Knez, M. (2001). The illusion of leadership: Misattribution of cause in coordination games, *Organization Science*, 12, 582-598.

Weber, J., Malhotra, D., & Murnighan, J. (2004). Normal acts of irrational trust: Motivated attributions and the trust development process, *Research in Organizational Behavior*, 26, 75-101.

8e6 Technologies (2008), 2008 Personal Internet Use Survey, http://www.8e6.com/US_UK_Survey_Results, (accessed 21 May 2010).

Table 1: Average Amounts Passed/Returned by Role, Condition, and Round Type

			First 10 Rounds				*All Rounds*			
			Monitoring		Non-Monitoring		Monitoring		Non-Monitoring	
	Condition		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Odd	Anticipated (n=44)	40%	3.82	(1.42)	2.94	(1.62)	3.83	(1.46)	2.90	(1.70)
	Unanticipated (n=38)	40%	3.06	(1.38)	2.92	(1.38)	3.24	(1.49)	3.00	(1.51)
	Anticipated (n=54)	80%	4.03	(1.19)	3.39	(1.73)	4.12	(1.17)	3.58	(1.52)
	Unanticipated (n=42)	80%	3.45	(1.17)	3.40	(1.65)	3.32	(1.30)	3.08	(1.90)
Even	Anticipated (n=44)	40%	7.00	(3.17)	3.14	(3.28)	7.05	(3.31)	3.17	(3.27)
	Unanticipated (n=38)	40%	4.54	(3.19)	4.33	(3.04)	4.88	(3.36)	4.57	(3.34)
	Anticipated (n=54)	80%	7.25	(3.02)	4.35	(4.11)	7.44	(3.01)	3.64	(3.65)
	Unanticipated (n=42)	80%	5.81	(2.94)	5.79	(3.59)	5.60	(3.11)	5.25	(3.92)

Table 2: Percent of the Total Endowment Even Players Returned to Odd

		First 10 Rounds				*All Rounds*			
		Monitoring		Non-Monitoring		Monitoring		Non-Monitoring	
Condition		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Anticipated	40%	41%	(0.16)	18%	(0.17)	40%	(0.16)	18%	(0.17)
Unanticipated	40%	32%	(0.14)	31%	(0.14)	32%	(0.14)	32%	(0.14)
Anticipated	80%	43%	(0.13)	29%	(0.22)	42%	(0.13)	24%	(0.21)
Unanticipated	80%	35%	(0.14)	34%	(0.20)	33%	(0.15)	30%	(0.21)

Note: These percentages exclude all cases in which Even players received 0 (and hence had no choice to make).

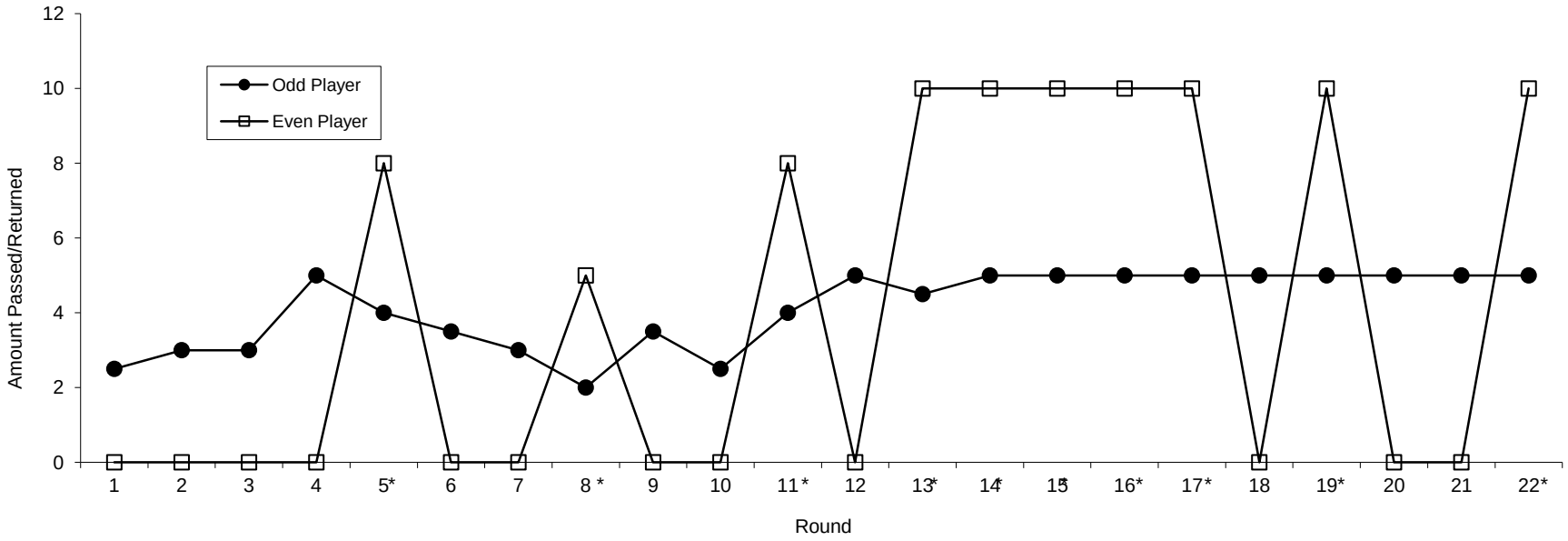
Table 3: Logistic Regression Analyses

	Model 1		Model 2		Model 3		Model 4	
	Parameter Estimate		Parameter Estimate		Parameter Estimate		Parameter Estimate	
Odd Player								
Intercept, α	-1.25 **	(.10)	-1.66 **	(.15)	-1.04 **	(.10)	-0.68 **	(.11)
Anticipated A, β_A	-0.44 **	(.09)	-0.06	(.25)	-0.52 *	(.21)	-0.82 **	(.10)
High Feedback H, β_H	0.48 **	(.12)	0.70 **	(.12)	0.57 **	(.13)	0.69 **	(.11)
Interaction F(r)*A, β_{FA}	0.91 **	(.10)	0.87 **	(.17)	0.82 **	(.12)	0.83 **	(.04)
# Trustworthy/ # Feedback, $\beta_{n(t)}$	0.96 **	(.10)	1.66 **	(.14)	0.29 **	(.10)	0.20 **	(.06)
Interaction A*n(t), $\beta_{An(t)}$	0.65 **	(.11)	-0.03	(.21)	0.68 **	(.19)	0.77 **	(.08)
Even Player								
Intercept, α	-1.03 **	(.14)	-1.13 **	(.13)	1.01 **	(.10)	1.48 **	(.12)
Anticipated A, β_A	-0.98 **	(.26)	-1.37 **	(.23)	-1.56 **	(.22)	-2.36 **	(.22)
High Feedback H, β_H	0.85 **	(.13)	0.80 **	(.12)	0.54 **	(.11)	0.30 **	(.10)
Interaction F(r)*A, β_{FA}	1.67 **	(.22)	2.13 **	(.16)	2.18 **	(.25)	2.77 **	(.22)
# Trustworthy/ # Feedback, $\beta_{n(t)}$	0.78 **	(.17)	1.04 **	(.11)	1.02 **	(.12)	0.20	(.15)
Interaction A*n(t), $\beta_{An(t)}$	0.33	(.24)	0.24	(.19)	-0.44 **	(.12)	0.48 *	(.23)
Log-likelihood	-847		-1308		-765		-1226	

Model 1: Odd Passes At Least 5 and Even Returns At Least 10 .
Every dyad played for at least 10 rounds. Results for this analysis include only the first ten rounds.
Model 2: Odd Passes At Least 5 and Even Returns At Least 10.
Results for this analysis include all of the rounds. Some dyads played games that had more rounds than others.
Model 3: Odd Passes At Least 5 and Even Returns Initial Endowment.
Every dyad played for at least 10 rounds. Results for this analysis include only the first ten rounds.
Model 4: Odd Passes At Least 5 and Even Returns Initial Endowment.
Results for this analysis include all of the rounds. Some dyads played games that had more rounds than others.

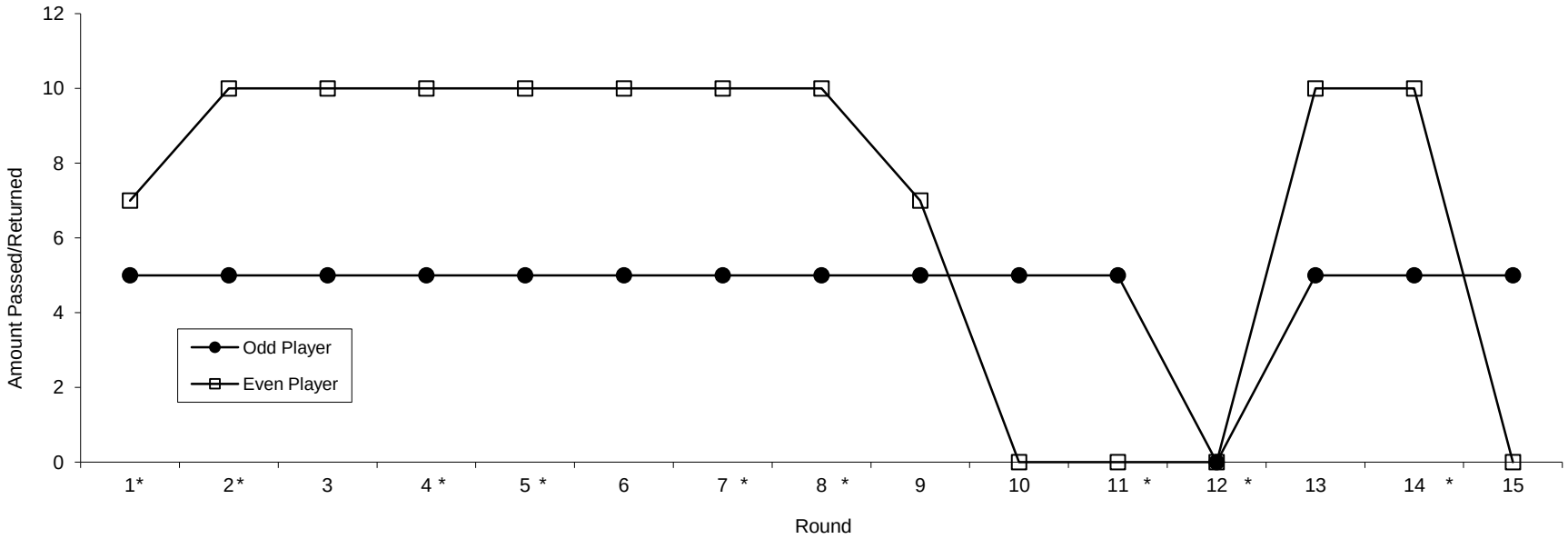
Figures

Figure 1: Example Dyad from the Anticipated 40% Condition



* Denotes a feedback round.

Figure 2: Example Dyad from the Unanticipated 80% Condition



* Denotes a feedback round.